(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **05249987 A**

(43) Date of publication of application: **28.09.93**

(51) Int. Cl **G10L 3/00**
G10L 9/00

(21) Application number: **04050327**

(22) Date of filing: **09.03.92**

(71) Applicant: **MATSUSHITA ELECTRIC IND CO LTD**

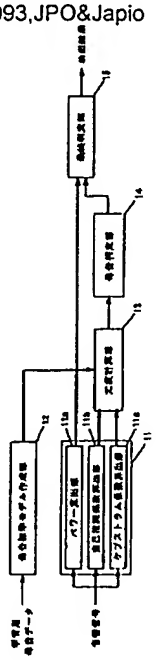
(72) Inventor: **NAKATO YOSHIHISA**
NORIMATSU TAKESHI

(54) VOICE DETECTING METHOD AND DEVICE

(57) Abstract:

PURPOSE: To automatically detect voice with high precision with a comparatively simple structure in a voice detecting device for detecting only the voice, which is used as the pretreatment of a voice recognizing device.

CONSTITUTION: A voice detecting device has a tolerance calculating part 13 for extracting a plurality of characteristic quantities from input signal every fixed time by a characteristic extracting part 11 and calculating the logarithmic tolerance with a vowel standard model formed by use of a number of learning data of vowel by a vowel standard model forming part 12: and a vowel judging part 14 for calculating a frame average logarithmic tolerance by collectively using the logarithmic tolerances for several frames and detecting vowels by comparison with a proper threshold. According to the number judged as vowels by the vowel judging part, it is judged by a final judging part 15 whether this section is a voice or not.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平5－249987

(43)公開日　平成５年(1993)９月28日

(51)Int.Cl.[5]		識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ	3/00	５１３　Ｂ	8842－5H		
	9/00	Ａ	8946－5H		

審査請求　未請求　請求項の数３(全　７　頁)

(21)出願番号　特願平4－50327

(22)出願日　平成４年(1992)３月９日

(71)出願人　000005821
松下電器産業株式会社
大阪府門真市大字門真1006番地

(72)発明者　中藤　良久
大阪府門真市大字門真1006番地　松下電器産業株式会社内

(72)発明者　則松　武志
大阪府門真市大字門真1006番地　松下電器産業株式会社内

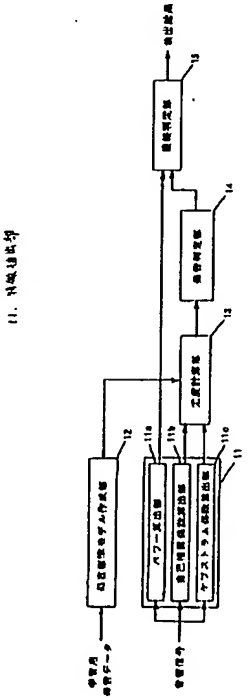
(74)代理人　弁理士　小鍛治　明　(外２名)

(54)【発明の名称】　音声検出方法および音声検出装置

(57)【要約】

【目的】　本発明は音声認識装置等の前処理として用いられる、音声のみを検出する音声検出装置に関するもので、比較的簡単な構成で自動的にしかも高精度に音声の検出を行うことができる音声検出装置を提供することを目的とする。

【構成】　入力信号から特徴抽出部１１にて一定時間毎に複数の特徴量を抽出し、母音標準モデル作成部１２であらかじめ多数の母音の学習データを用いて作成した母音標準モデルとの対数尤度を計算する尤度計算部１３と、数フレーム分の対数尤度を一括して用いてフレーム平均対数尤度を計算し、適当な閾値と比較することで母音を検出する母音判定部１４と、母音判定部により母音と判定された個数により、最終判定部１５にてその区間が音声か否かを判定する。



1

【特許請求の範囲】

【請求項１】　入力信号からフレーム単位（一定時間毎）に抽出した音声を特徴付ける１次以上の自己相関係数もしくは偏自己相関関数と１次以上のケプストラム係数もしくはメルケプストラム係数のうち少なくとも１つの特徴量を用いて、その数フレーム分を一括して用いることにより母音の存在確率を求め母音検出を行い、音声のみを検出することを特徴とする音声検出方法。

【請求項２】　あらかじめ多数の母音の学習データについてフレーム単位に抽出した音声を特徴付ける１次以上の自己相関係数もしくは偏自己相関関数と１次以上のケプストラム係数もしくはメルケプストラム係数のうち少なくとも１つの特徴量を用いて、数フレーム分を一括して用いて母音毎に標準モデルの作成を行い、前記母音標準モデルを用いて音声を検出することを特徴とする音声検出方法。

【請求項３】　入力信号から一定時間毎に音声を特徴付ける１次以上の自己相関係数もしくは偏自己相関関数と１次以上のケプストラム係数もしくはメルケプストラム係数を抽出する特徴抽出部と、あらかじめ多数の母音の学習データについて前記特徴抽出部で抽出した特徴量を用いて母音毎の平均値と共分散行列を算出し、母音毎の標準モデルを作成する母音標準モデル作成部と、入力信号からフレーム単位に前記特徴抽出部で抽出した１次以上の自己相関係数もしくは偏自己相関関数と１次のケプストラム係数もしくはメルケプストラム係数のうち少なくとも１つ以上の特徴量について、前記母音標準モデル作成部にて作成した各母音標準モデルとの対数尤度を計算する尤度計算部と、母音検出しようとするフレームとその前後数フレームにおいて前記尤度計算部にて計算された対数尤度を用いて各母音毎にフレーム平均対数尤度を計算し、ある適当な閾値と比較することで母音かそれ以外かを判定する母音判定部と、パワーの一定レベル以上の入力信号の塊について前記母音判定部により、いずれかの母音と判定された母音サンプルの個数がある適当な閾値以上のときにその塊を音声と判定する最終判定部とを備えたことを特徴とする音声検出装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、非定常雑音の存在する実環境下において、音声認識装置の前処理等で使われる、音声のみを検出する音声検出方法および音声検出装置に関する。

【０００２】

【従来の技術】音声認識等の音声処理を行う装置では、音声以外の非定常雑音が入力され誤って音声と判断されると誤認識を生じる。そこで、入力された信号が正確に音声であるかどうかを判定できる音声検出装置が必要とされる。

【０００３】従来の音声検出装置では、処理の簡素化のための入力信号のパワー値が閾値よりも大きい部分を音声と判断する方法が一般的に行われる。しかし音声認識の行われる実環境で使用することを考えると、紙などの資料をめくる音や、息吹きなどのマイクロフォンの振動によって起こるノイズ、あるいは動物の鳴き声等の音声以外のパワーの大きな様々な音が入力される可能性があり、パワーだけでは音声の検出はできない。

2

【０００４】そこで、パワー以外の複数の音声の特徴量を用いて入力信号が音声であるか非音声であるかの判定をする方法が幾つか提案されている。例えば、「実環境下での音声／非音声の判別」（石田明・小畑秀文、日本音響学会誌７巻１２号（１９９１））による方法がある。これは、日常の実験室やオフィスなどで発生する種々の非定常雑音と音声とを区別するのに有効な音響的特徴量を用いて、実環境下での音声／非音声の判別を行っている。具体的には、音声中のパワーの大きい部分において、母音と見なせる部分がどの程度存在するかによって、音声／非音声の判別を行っており、用いる音響的特徴量としては、

（ａ）周期性
（ｂ）ピッチ周波数
（ｃ）最適線形予測次数
（ｄ）５母音との距離
（ｅ）ホルマントの鋭さ

の５種類の特徴量を求め、各特徴量毎に上限値あるいは下限値を決定し、その大小関係により音声と非音声を判別する。

【０００５】

【発明が解決しようとする課題】しかしながら上記の音声／非音声判別装置では、音声中の各母音の特徴に基づいた特徴量は使用されておらず、音声の各母音の検出に適した母音毎の標準モデルを用いる方法による高精度な音声検出方式が必要とされる。また、各特徴量毎に上限値あるいは下限値を決定し、その大小関係により音声／非音声の判別を行う方法では、特徴量の数が増えた場合特徴量毎に閾値を設定することが困難であると同時に、定常雑音が付加された場合などのような特徴量の値の変動に対して頑健であるとは言えない。さらに、音声、特に母音は１分析フレーム毎に母音性を判定するより、数フレーム分を１塊に考えて判定する方がより信頼性がある判定法であるといえる。

【０００６】本発明は、上記の課題を解決するもので、音声認識等の音声信号処理に適した高性能な音声検出装置を提供することを目的とする。本発明は、音声の各母音の検出に適した母音毎の標準モデルを用いることで、音声の各母音の検出に基づいた音声検出装置を提供する。さらに、音声であるかそれ以外であるかを表した特徴量を総合的に判定するための評価値として、母音性を判定するのに有用と考えられる数フレーム分を１塊に考えて算出されるフレーム平均対数尤度を用いることで、

3

閾値の設定が比較的容易であると同時に、定常雑音が付加された場合などのような特徴量の値の変動に対してある程度頑健性を持った、比較的簡単で高性能な音声検出装置を提供することを目的とする。

【０００７】

【課題を解決するための手段】本発明は上記課題を解決するために、入力信号からフレーム単位（一定時間毎）に抽出した音声を特徴付ける１次以上の自己相関係数と１次以上のケプストラム係数のうち少なくとも１つの特徴量を用いて、その数フレーム分を一括して用いることにより母音の存在確率を求めて母音検出を行い、これにより音声のみを検出することを特徴とするものである。

【０００８】また、本発明は、あらかじめ多数の母音の学習データについてフレーム単位に抽出した音声を特徴付ける１次以上の自己相関係数と１次以上のケプストラム係数のうち少なくとも１つの特徴量を用いて、数フレーム分を一括して用いて母音毎に標準モデルの作成を行い、前記母音標準モデルを用いて音声を検出することを特徴とするものである。

【０００９】さらに、本発明の音声検出装置は、母音を検出することを主眼として、入力信号の一定時間毎の１次以上の自己相関係数、１次以上のケプストラム係数等の複数の音声の特徴量を抽出する特徴量抽出部と、あらかじめ多数の母音の学習データについて前記特徴抽出部で抽出した特徴量を用いて母音毎の平均値と共分散行列を算出し、母音毎の標準モデルを作成する母音標準モデル作成部と、入力信号からフレーム単位に前記特徴抽出部で抽出した１次以上の自己相関係数と１次のケプストラム係数のうち少なくとも１つの特徴量について、前記母音標準モデル作成部にて作成した各母音標準モデルとの対数尤度を計算する尤度計算部と、前記尤度計算部にて計算された前後数フレームについて各母音毎にフレーム平均対数尤度を計算し、ある適当な閾値とを比較することで母音かそれ以外かを判定する母音判定部と、パワーの一定レベル以上の入力信号の塊について前記母音判定部によりいずれかの母音と判定された母音サンプルの個数の割合がある適当なしきい値以上のときにその塊を音声と判定する最終判定部とを備えたものである。

【００１０】

【作用】本発明は、上記した構成により、音声中の各音韻の特徴に基づく母音検出に適した特徴量を用い、あらかじめ信頼性の高い多数の母音データを用いて母音毎に母音標準モデルを作成し、数フレーム分を一括して統計的手法により母音の検出を行い、音声のみを検出することで、高性能な音声検出が可能となる。

【００１１】

【実施例】以下本発明の一実施例について説明する。

（図１）は本発明の一実施例の全体構成を示すブロック構成図である。（図１）において、１１は音声検出のための複数の特徴量を抽出する特徴抽出部で、１フレーム

4

（一定時間）毎のパワーを計算するパワー計算部１１ａと、１フレーム毎の１次および７次の自己相関係数を算出する自己相関係数算出部１１ｂと、１フレーム毎の１次および３次のケプストラム係数を算出するケプストラム係数算出部１１ｃとから構成される。これらの特徴量は入力信号の母音性を検出するために用いられる。

【００１２】次に、１２はあらかじめ多数の母音の学習データについて特徴抽出部１１で抽出した特徴量を用いて母音毎の平均値と共分散行列を算出し、母音毎の標準モデルを作成する母音標準モデル作成部である。１３は特徴抽出部１１から出力されるフレーム毎の入力信号の１次および７次の自己相関係数と１次および３次のケプストラム係数について、母音標準モデル作成部１２にて作成した各母音標準モデルとの対数尤度を計算する尤度計算部であり、１４は尤度計算部１３にて計算に用いたフレームの前後数フレームにおいて、尤度計算部１３で同様に計算された対数尤度を用いて、各母音毎にフレーム平均対数尤度を計算し、ある適当な閾値と比較することでその入力信号数フレームが母音であるかどうかを判定する母音判定部である。１５はパワーの一定レベル以上の入力信号の塊について母音判定部１４によりいずれかの母音と判定されたフレームの個数がある適当な閾値以上のときにその塊を音声と判定する最終判定部である。

【００１３】以下、本発明の一実施例について（図１）のブロック構成図を参照しながら詳細に説明する。音響信号がマイクロホンを通して入力されると、特徴抽出部１１でまず複数の特徴量が抽出される。パワー計算部１１ａでは、一定時間毎のパワー値が例えば（数１）で算出される。一定の時間間隔は、ここでは例えばサンプリング周波数を１０ＫＨｚとして、２００点（２０ｍｓ）とし、この時間単位をフレームと呼ぶ。

【００１４】

【数１】

$$P_i = \sum_{k=1}^{200} S_k * S_k$$

【００１５】ここで、Ｐｉはフレームｉでのパワー値、Ｓｋはフレーム内の入力信号のサンプル値を示す。このパワー値は発声条件の違いによるパワーの違いを統一して扱えるように、パワーの大きな区間内の最大値、最小値間を例えば０から１までの値に正規化して用いる。自己相関係数算出部１１ｂではフレーム毎に１次の自己相関係数Ａｉ(1)が（数２）、７次の自己相関係数Ａｉ(7)が、（数３）で算出される。

【００１６】

【数２】

$$A_i(1) = \sum_{k=1}^{200} S_k * S_{k+1}$$

【００１７】

【数３】

5

$$A_i(7)=\sum_{k=1}^{200} S_k * S_{k+7}$$

【0018】さらにAi(1)、Ai(7)は0次の自己相関係数で正規化される。ケプストラム係数算出部11cでは、フレームiでの1次および3次のケプストラム係数Ci(1)、Ci(3)が線形予測分析により求められる。

【0019】母音標準モデル作成部12では、あらかじめ多数の音声データの母音部分について特徴抽出部11で得られる特徴量を抽出しておき、これらの特徴量を用いて次の方法により母音毎の平均値と共分散行列を算出し、母音毎の標準モデルを作成する。すなわち、母音データとしては母音kの学習用データ y_N（データ数N）を用い、y_N がm次元の多次元正規分布に従うと仮定できる場合、その平均値 μ_k と共分散行列 Σ_k を（数4），（数5）のように計算にて求めることができる。

【0020】

【数4】

$$\mu_k = \frac{1}{N}\sum_{n=1}^{N} y_n$$

【0021】

【数5】

$$\Sigma_k = \frac{1}{N}\sum_{n=1}^{N}(y_n-\mu_k)(y_n-\mu_k)'$$

6

*【0022】ただし、tは転値を示す。これにより、母音毎の標準モデルのモデル形状（平均値 μ_k、及び分散 Σ_k）が求められる。ただし、y_N、μ_kはm次元のベクトル（m次元の特徴パラメータ）であり、Σ_kはm×m次元のマトリックスである。母音データとしては例えば、ある標準話者の母音kの学習用データとして母音部分を切り出し、母音中心フレーム±2フレームのデータを用いればよい。また、複数の話者のデータを用いることで、話者の発声の変動に強い標準パターンを作成することができる。

【0023】尤度計算部13は、特徴抽出部11から出力されるフレーム毎の入力信号のいくつかの特徴パラメータについて、母音標準モデル作成部12にて作成した各母音標準モデルとの対数尤度を計算する部分である。母音検出に用いる距離尺度は、使用する各特徴パラメータの分布を多次元正規分布と仮定した場合の統計的距離尺度である。母音毎の標準モデルkに対する、iフレーム目の入力ベクトル（スペクトル）xi の対数尤度Likは、（数6）で計算される。

【0024】

【数6】

*

$$L_{ik}=-\frac{1}{2}(x_i-\mu_k)'\Sigma_k^{-1}(x_i-\mu_k)-\frac{1}{2}\ln|\Sigma_k|+C$$

【0025】ただし、xiはm次元のベクトル（m次元の特徴パラメータ）であり、tは転値、−1は逆行列、Cは定数を示す。

【0026】母音判定部14は、母音の時間的な継続性を表現するため、検出しようとする目的フレームの前後Nフレーム（これをセグメントという）を用いて、母音判定を行う。各母音毎に尤度計算部13にて計算された対数尤度Likを用いて、次の条件式（数7）を満たせばそのセグメントは母音であるとみなす。

【0027】

【数7】

$$\frac{1}{2*N+1}\sum_{i=-N}^{N} L_{ik} \geqq L_{kTH}$$

【0028】ただし、L_{kTH} は母音標準モデルkに関する判別閾値（フレーム平均対数尤度の閾値）である。

【0029】このように、各特徴パラメータの影響を効果的に、しかも総合的に判定できる評価値（対数尤度）を用いることで、各特徴パラメータ毎に閾値を設定する方法よりも、定常雑音が付加された場合などのような特徴量の値の変動に対して頑健なシステムが構築できる。また、多くの閾値をヒューリスティックな方法により決定する必要がない利点がある。さらに、音響信号数フレーム分を1塊に考えて判定することで、母音などのような継続的な音声に対してより有効な判定法となっている。

【0030】最終判定部15は、パワーの一定レベル以上の入力信号の塊についての母音サンプルの数がある適当なしきい値以上のときにその塊を音声と判定する最終判定部である。最終判定部15では、まずパワー計算部11aで得られたパワー値系列からあらかじめ定めたパワーしきい値を決められた長さ以上越える区間を音声候補区間として検出する。この音声候補区間内において、母音判定部14により母音kと判定されたセグメントの個数を計数し、母音kと判定された母音セグメントの数を C_k、あらかじめ定めた区間内の母音セグメント数のしきい値 M_k とするとき、（数8）の条件を満たすならば、この音声候補区間は音声であると判定する。これを全ての母音について行う。

【0031】

【数8】

$$C_k \geqq M_k$$

【0032】以下に、実際に本方法により実験した結果を示す。（表1）に、本手法で用いた4つの特徴パラメ

7

ータを示す。これらの特徴パラメータは予備実験の結果、音声と他の非定常雑音との分離が比較的良く、またＬＰＣケプストラム係数の算出過程において容易に得られるパラメータである。まず、１次の正規化自己相関係数及び１次の線形予測係数は有声／無声の判別に適したパラメータであり、７次の正規化自己相関係数は低周波性の雑音を区別するのに適したパラメータである。また、３次のＬＰＣケプストラム係数は、５母音の中でも／ｉ／に特徴的な性質を示すパラメータである。

【００３３】

【表１】

・１次の自己相関係数
・１次の線形予測係数
・７次の自己相関係数
・３次のＬＰＣケプストラム係数

8

＊【００３４】音声データは、男性１０名の発声した日本語２００単語である。標準モデルの作成には、標準話者の発声した各母音の音韻中心±２フレームを使用した。但し、計算効率を考えて、各パラメータ間の相関はないとし、共分散行列の対角成分のみを計算に用いた。雑音データとしては、（表２）に示す５雑音グループ（約９００サンプル）の非定常雑音を用いた。また、分析条件を（表３）に示す。

【００３５】

【表２】

＊

非定常雑音グループ	サンプル数
紙の音	２８５
手や物で机を叩く音	２１５
ガラス・食器の触れ合う音	１００
マイクに触れる・叩く音	１８４
息吹き・咳払い	１１３

【００３６】

【表３】

サンプリング周波数	10[kHz]
分析窓	20[ms]ﾊﾐﾝｸﾞ窓
フレーム周期	10[ms]
プリエンファシス	$1-0.9z^{-1}$
ＬＰＣ分析次数	12次

【００３７】男性話者５名分の母音データから標準モデルの作成を行い、標準話者を含む１０名の話者に関しての音声検出実験及び（表２）の非定常雑音の除去実験を行った。（図２）は、母音セグメント長を１フレームから１１フレームまで変化させたときの、音声検出率と雑音誤検出率の関係を示したものである。判別閾値を適当に変化させることで、検出性能の最適値を求めることができるが、５フレーム以上ではほとんど判別性能に差はない。結局、母音セグメント長７フレームで判別閾値＝-1.2のとき、音声検出率９９．３％（雑音誤検出率９．０％）が得られた。

【００３８】次に、本手法の定常騒音下での検出性能を評価するために、白色雑音を付加したときのＳ／Ｎ比と検出率との関係を調べた。（図３）は、母音セグメント長を７フレームに固定したときの、各Ｓ／Ｎ比に対する音声検出率と雑音誤検出率との関係を示したものである。その結果、検出性能はＳ／Ｎ比が１２ｄＢまでほとんど影響を受けていない。

【００３９】以上のように本実施例の音声検出装置によれば、入力信号から一定時間毎の音声の複数の特徴量を抽出する特徴量抽出部１１と、あらかじめ多数の母音に関する学習データについてフレーム単位で抽出した前記特徴量を用いて母音毎の平均値と共分散行列を算出し、母音毎の標準モデルを作成する母音標準モデル作成部１２と、入力信号から得られた複数の音声の特徴量と母音標準モデル作成部にて作成した各母音標準モデルとの対数尤度を計算する尤度計算部１３と、母音判定を行うフレームの前後数フレーム分の対数尤度を用いて、各母音毎にフレーム平均対数尤度を計算し、ある適当なしきい値とを比較することでその入力信号数フレーム分が母音であるかどうかを判定する母音判定部１４と、パワーの一定レベル以上の入力信号の塊について母音判定部１４によりいずれかの母音と判定された母音サンプルの個数がある適当なしきい値以上のときにその塊を音声と判定

9

する最終判定部１５とを具備して構成することにより、比較的簡単な構成で様々な音響信号の中の音声を正確に判定することができる音声検出装置を提供することができる。

【００４０】なお、上記の実施例においては、特徴抽出部において入力信号の母音性を検出するための特徴量として自己相関係数とケプストラム係数を用いた例で説明したが、これに限定されず、偏自己相関関数やメルケプストラム係数などを用いてもかまわない。

【００４１】

【発明の効果】以上の実施例から明らかなように本発明によれば、音声を特徴付ける複数の特徴量を抽出し、多数の学習用母音データを用いて母音標準モデルを作成しておき、入力信号から得られた複数の特徴量と母音標準モデルから得られる対数尤度を計算し、数フレーム分を一括して母音の検出を行って音声を検出するように構成しているので、比較的簡単な構成で入力信号が音声かそれ以外かを正確に判定することができる音声検出装置を

10

提供することができる。

【図面の簡単な説明】

【図１】本発明の一実施例の音声検出装置の全体構成を示すブロック図

【図２】母音セグメント長の影響による音声検出率と雑音誤検出率の関係を示す図

【図３】Ｓ／Ｎ比の影響による音声検出率と雑音誤検出率の関係を示す図

【符号の説明】

10　１１　特徴抽出部

１１ａ　パワー算出部

１１ｂ　自己相関係数算出部

１１ｃ　ケプストラム係数算出部

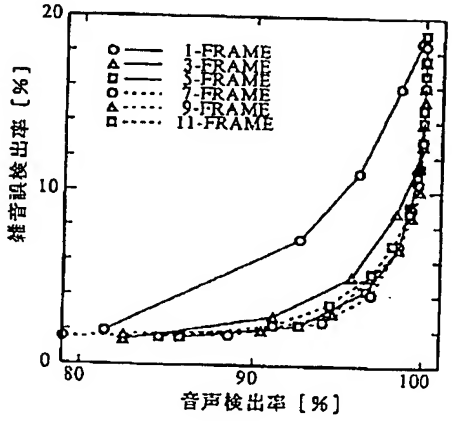
１２　母音標準モデル作成部

１３　尤度計算部

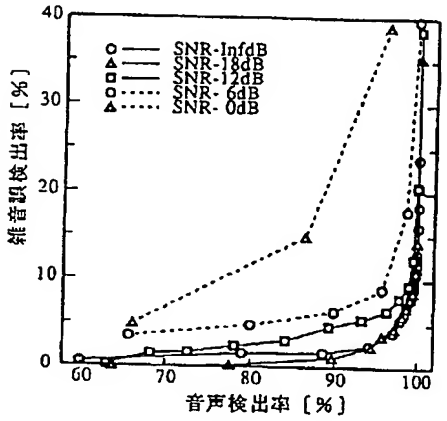
１４　母音判定部

１５　最終判定部

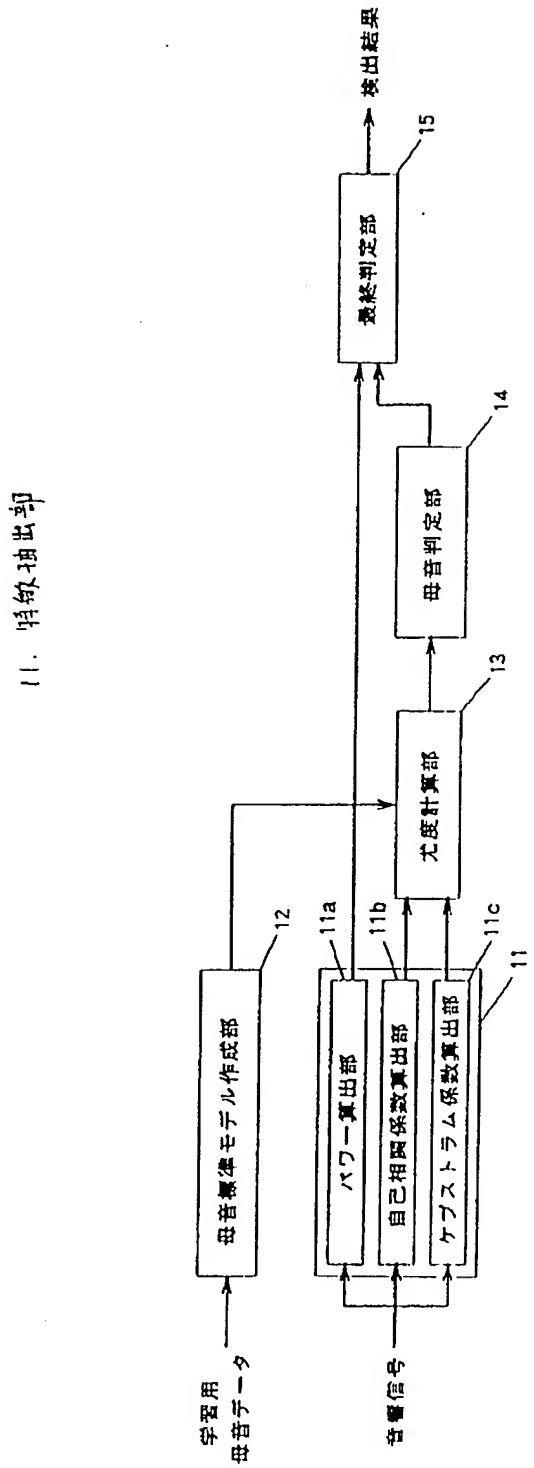
【図２】



【図３】



【図1】


検出結果
15
最終判定部
14
母音判定部
13
尤度計算部
11a
11b
11c
11
12
パワー算出部
自己相関係数算出部
ケプストラム係数算出部
母音標準モデル作成部
学習用
母音データ
音響信号
11. 特徴抽出部